AD________________

Award Number: DAMD17-02-1-0558

TITLE: Statistical Inference for Quality-Adjusted Survival Time

PRINCIPAL INVESTIGATOR: Hongwei Zhao, Sc.D.
David Oakes, Ph.D.

CONTRACTING ORGANIZATION: University of Rochester
Rochester, New York 14627

REPORT DATE: August 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2004	Annual Summary (15 Jul 03 - 14 Jul 04)

4. TITLE AND SUBTITLE
Statistical Inference for Quality-Adjusted Survival Time

5. FUNDING NUMBERS
DAMD17-02-1-0558

6. AUTHOR(S)
Hongwei Zhao, Sc.D.
David Oakes, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Rochester
Rochester, New York 14627

E-Mail: zhao@bst.rochester.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

In evaluations of breast cancer therapies, the patients' quality of life is receiving more and more attention. It is desirable that a treatment not only prolongs the overall survival life, but also improves the quality of life (QOL). Quality-adjusted lifetime (QAL) is a measure that combines both the quality and the quantity of a person's lifetime. The goal of my research is to study how to draw inference about QAL in the presence of censoring. For the past year I have made some progress on the regression problem of QAL. I am confident that the research problem I have been working on is of great importance to breast cancer studies.

14. SUBJECT TERMS
Quality of life, censored data, survival analysis

15. NUMBER OF PAGES
8

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

	Page Number
Cover	
SF 298	
Table of Contents	
Introduction	1
Body	1
Key Research Accomplishments	3
Reportable Outcomes	4
Conclusions	4
References	5

Annual Summary Report

Statistical Inference for Quality-Adjusted Survival Time

Hongwei Zhao, Sc.D

a. Introduction

In evaluations of breast cancer therapies, the patients' quality of life is receiving more and more attention. It is desirable that a treatment not only prolongs the overall survival life, but also improves the quality of life (QOL). Quality-adjusted lifetime (QAL) is a measure that combines both the quality and the quantity of a person's lifetime. First proposed by Gelber, Gelman and Goldhirsch (1989), QAL is simply an integration of survival time weighted by a utility coefficient ranging from 0 (poor health) to 1 (perfect health). In a typical clinical trial setting, patients are enrolled over time, and the study ends before observation of the endpoints for all patients. Therefore, the data are right censored. The goal of my research is to study how to draw inference about QAL in the presence of censoring.

For this past year, I have concentrated on the regression problem for the QAL, since this topic has received great interest lately.

b. Body

b.1. The Setting

For the ith individual in the study, let's define $V_i(t), t \geq 0$, as a continuous time stochastic process representing the patient's health history process, T_i as the survival time. U is a utility function, which is assumed to be known or can be specified. The ith individual's quality adjusted lifetime, denoted as Q_i, is equal to

$$Q_i = \int_0^{T_i} U\{V_i(t)\}dt.$$

Let Z_i denote the $p+1$ vector of covariates associated with the ith individual, and C_i denote the censoring variable. We assume that the censoring variable is independent of the health history, conditional on the covariates.

Due to the presence of censoring, we cannot make inference on QAL over the entire health history. We can only consider the QAL accumulated within a time L, where L is a time limit up to which we have reasonable amount of data. Our inference therefore will depend on the choice of L. Consequently, the survival time of a patient will be truncated at L, that is, $T^L = \min(T, L)$. For ease of notation, however, we still use T instead of T^L in subsequent development of the theory.

We observe the following data:

$$X_i = \min(T_i, C_i), \Delta_i = I(T_i < C_i), V_i(t), 0 \leq t \leq X_i,$$

$$Q_i = \int_0^{X_i} U\{V_i(t)\}dt, Z_i, \quad i = 1, \cdots, n.$$

We would like to investigate which covariates affect the quality adjusted survival time.

b.2. Regression models for the mean QAL

We assume a general model for the mean QAL:

$$E(Q_i) = f(\beta^T Z_i),$$

where β is a $p+1$ vector of coefficients, f(.) is a known function.

If we can observe Q_i for every subject, then a consistent estimating equation for solving β is

$$\sum_{i=1}^{n} h(Z_i)\{Q_i - f(\beta^T Z_i)\} = 0,$$

where $h(Z_i)$ is any function of Z_i. Without censoring, it has been shown that the best choice of the function $h(Z_i)$ is $\partial f(\beta^T Z_i)/\partial\beta$.

When censoring is present, we can not observe Q_i for everybody. Instead, we use the idea of inverse probability weighting, and propose the following estimating equation:

$$S_n(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{\widehat{K}(T_i)} h(Z_i)\{Q_i - f(\beta^T Z_i)\} = 0. \tag{1}$$

It is easy to show that this estimating equation will produce consistent estimators for β.

b.3. Efficiency Issue

In the above estimating equation, only the data on QAL for the people who have failures are used, the QAL for censored subjects are ignored. Hence the above estimating equation is not efficient.

Using the counting process representation similar to Zhao and Tsiatis (1997), we can write:

$$\begin{aligned} & n^{\frac{1}{2}} S_n(\beta) \\ \approx \;& n^{-\frac{1}{2}} \sum_{i=1}^{n} B_i - n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^{\infty} \frac{dM_i^c(u)}{K(u)} \{B_i - G(B, u)\}, \end{aligned} \tag{2}$$

where

$$B_i = h(Z_i)\{Q_i - f(\beta^T Z_i)\},$$

and for any random variable X, function $G(X, u)$ is defined as

$$G(X, u) = \frac{\mathrm{E}\{X_i I(T_i \geq u)\}}{S_T(u)}.$$

Here, $S_T(u) = \mathrm{pr}(T > u)$.

Using the theory of Robins and Rotnitzky (1992), and Robins, Rotnitzky and Zhao (1994), a class of estimating equations can be written as

$$S_n(e, \beta) = S_n(\beta) + \sum_{i=1}^{n} \int_0^\infty \frac{dM_i^c(u)}{K(u)} [e\{V_i^H(u)\} - G(e, u)],$$

where $e\{V_i^H(u)\}$ is a functional of the health history process.

The key findings are:

(1) For any function $h(Z_i)$ chosen, the most efficient estimating equation is one with

$$e_{eff}\{V_i^H(u)\} = \mathrm{E}\{B_i | V_i^H(u)\}.$$

(2) The optimal function $h(Z_i)$ is

$$h_{opt}(Z) = E\{\partial / \beta R(\beta) | Z\} Var\{R(\beta) | Z\}^{-1},$$

where $R_i(\beta) = \frac{\Delta_i}{K(T_i)} D_i + \int_0^\infty \frac{dM_i^c(u)}{K(u)} [\mathrm{E}\{D_i | V_i^H(u)\} - G(e, u)]$, $D_i = Q_i - f(\beta^T Z_i)$.

Although we derived the formula for the most efficient estimating equation, this estimating equation depends on complicated health history which is not known to us and has to be estimated.

My research has been focusing on finding a functional form of $e\{V_i^H(u)\}$ that is as close to $e_{eff}\{V_i^H(u)\}$ as possible, and on estimating the $h_{opt}(Z)$. I have been conducting simulation studies to test these methods. I will apply the methods to a real breast cancer example.

c. Key Research Accomplishment

I have derived the most efficient estimating equation for the regression problem of the mean QAL.

d. Reportable outcomes

I have been invited to give a short course entitled "Statistical Inference of Quality Adjusted Lifetime" in International Chinese Statistical Association (ICSA) 2005 Applied Statistics Symposium at Washington, DC on Saturday, June 19, 2005.

e. Conclusions

I believe the research problem I have been working on is of great importance to breast cancer studies. For the past year I have made some progress on the regression problem of QAL. I am confident that I will obtain some good results by the end of the grant period.

References

Gelber, R.D., Gelman, R.S. and Goldhirsch, A. (1989). A quality-of-life oriented endpoint for comparing therapies. *Biometrics* **45**, 781-795.

Robins, J.M. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidemiology-Methodological Issues*, eds. Jewell, N, Dietz, K, and Farewell, V. Boston: Birkhäuser, 297-331.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89**, 846-866.

Zhao, H. and Tsiatis, A.A. (1997). A consistent estimator for the distribution of quality adjusted survival time. *Biometrika* **84**, 339-348.